

```

1 ATGGGGGAGA TGCAGGGCGC GCTGGCCAGA GCCCGGCTCG AGTCCCTGCT
51 GCGGCCCCGC CACAAAAAGA GGGCCGAGGC GCAGAAAAGG AGCGAGTCCT
101 TCCTGCTGAG CGGACTGGCT TTCATGAAGC AGAGGAGGAT GGGTCTGAAC
151 GACTTTATTC AGAAGATTGC CAATAACTCC TATGCATGCA AACACCCTGA
201 AGTTCAGTCC ATCTTGAAGA TCTCCCAACC TCAGGAGCCT GAGCTTATGA
251 ATGCCAACCC TTCTCCTCCA CCAAGTCCTT CTCAGCAAAT CAACCTTGGC
301 CCGTCGTCCA ATCCTCATGC TAAACCATCT GACTTTCACT TCTTGAAAGT
351 GATCGGAAAG GGCAGTTTGT GAAAGGTTCT TCTAGCAAGA CACAAGGCAG
401 AAGAAGTGTT CTATGCAGTC AAAGTTTAC AGAAGAAAGC AATCCTGAAA
451 AAGAAAGAGG AGAAGCATAT TATGTCGGAG CGGAATGTTT TGTTGAAGAA
501 TGTGAAGCAC CCTTTCCTGG TGGGCCTTCA CTTCTCTTTC CAGACTGCTG
551 ACAAATTGTA CTTTGTCTTA GACTACATTA ATGGTGGAGA GTTGTCTTAC
601 CATCTCCAGA GGGAAACGCTG CTTCTGGA CCACGGGCTC GTTCTATGC
651 TGCTGAAATA GCCAGTGCCT TGGGCTACCT GCATTCACCT AACATCGTTT
701 ATAGAGACTT AAAACCAGAG AATATTTTGC TAGATTCACA GGGACACATT
751 GTCCTTACTG ACTTCGGACT CTGCAAGGAG AACATTGAAC ACAACAGCAC
801 AACATCCACC TTCTGTGGCA CGCCGGAGTA TCTCGCACCT GAGGTGCTTC
851 ATAAGCAGCC TTATGACAGG ACTGTGGACT GGTGGTGCCT GGGAGCTGTC
901 TTGTATGAGA TGCTGTATGG CCTGCCGCCT TTTTATAGCC GAAACACAGC
951 TGAAATGTAC GACAACATTC TGAACAAGCC TCTCCAGCTG AAACCAAATA
1001 TTACAAATTC CGCAAGACAC CTCCTGGAGG GCCTCCTGCA GAAGGACAGG
1051 ACAAAGCGGC TCGGGGCCAA GGATGACTTC ATGGAGATTA AGAGTCATGT
1101 CTTCTTCTCC TTAATTAACT GGGATGATCT CATTAAATAAG AAGATTACTC
1151 CCCCTTTTAA CCCAAATGTG AGTGGGCCCCA ACGACCTACG GCACTTTGAC
1201 CCCGAGTTTA CCGAAGAGCC TGTCCCCAAC TCCATTGGCA AGTCCCCTGA
1251 CAGCGTCCTC GTCACAGCCA GCGTCAAGGA AGCTGCCGAG GCTTTCCTAG
1301 GCTTTTCCTA TGCGCCTCCC ACGGACTCTT TCCTCTGA (SEQ ID NO:1)

```

FEATURES:

Start Codon: 1

Stop Codon: 1336

Homologous proteins:

Top 10 BLAST Hits

		Score	E
CRA 67000082668077	gi 14756346 /def=ref XP_037046.1 (XM...	843	0.0
CRA 18000005074572	gi 5032091 /def=ref NP_005618.1 (NM...	841	0.0
CRA 18000004907445	gi 477098 /def=pir A48094 serum and ...	829	0.0
CRA 150000165029864	gi 13431833 /def=sp Q9XT18 SGK_RABIT...	826	0.0
CRA 18000005246968	gi 6755490 /def=ref NP_035491.1 (NM...	824	0.0
CRA 18000004937507	gi 9507093 /def=ref NP_062105.1 (NM...	822	0.0
CRA 18000005171986	gi 3688803 /def=gb AAC62398.1 (AF057...	776	0.0
CRA 18000005144813	gi 3116066 /def=emb CAA11528.1 (AJ22...	711	0.0
CRA 18000005144812	gi 3116064 /def=emb CAA11527.1 (AJ22...	709	0.0
CRA 335001098677651	gi 11321321 /def=gb AAG34115.1 AF312...	579	e-164

FIGURE 1A

Blast hits to dbEST:

CRA Number	gi Number	Score	Expect
CRA 63000075619018	gi 14816226	1562	0.0
CRA 225000014874111	gi 18504859	1503	0.0
CRA 225000001750119	gi 15756574	1441	0.0
CRA 159000009754651	gi 13582978	1417	0.0
CRA 158000041295522	gi 10994817	1407	0.0
CRA 58000099052833	gi 12793499	1402	0.0
CRA 66000078090204	gi 15017913	1402	0.0
CRA 225000014943770	gi 18509828	1392	0.0
CRA 63000075528266	gi 14809983	1368	0.0
CRA 78000169320891	gi 14073071	1364	0.0
CRA 335001063053989	gi 10937149	1296	0.0
CRA 78000106804089	gi 10390589	1285	0.0
CRA 118000029469319	gi 10933084	1259	0.0
CRA 11000545765847	gi 9176035	1239	0.0
CRA 222000003126349	gi 16520713	1225	0.0
CRA 158000041316197	gi 10996884	1223	0.0
CRA 55000120105106	gi 13994615	1197	0.0
CRA 222000003339745	gi 16529516	1191	0.0
CRA 78000169332857	gi 14074159	1183	0.0
CRA 224000000151825	gi 15161415	1181	0.0
CRA 158000041310407	gi 10996305	1174	0.0
CRA 224000004588220	gi 15950481	1172	0.0
CRA 98000052723591	gi 13976289	1158	0.0
CRA 222000000718505	gi 15248907	1156	0.0
CRA 78000105668173	gi 10352924	1152	0.0
CRA 164000029925315	gi 11001544	1146	0.0
CRA 164000029914485	gi 11000461	1132	0.0
CRA 107000020244468	gi 9330126	1132	0.0
CRA 55000120090850	gi 13993063	1132	0.0
CRA 158000041316657	gi 10996930	1122	0.0
CRA 78000106872608	gi 10398650	1122	0.0
CRA 165000138685182	gi 14292101	1100	0.0
CRA 164000029922335	gi 11001246	1094	0.0
CRA 335001063029672	gi 10947466	1094	0.0
CRA 113000083010569	gi 12041410	1086	0.0
CRA 222000001551530	gi 15446801	1070	0.0
CRA 78000105702017	gi 10357421	1053	0.0
CRA 107000020410692	gi 9345242	1049	0.0
CRA 107000020408053	gi 9345002	1049	0.0
CRA 196000006451565	gi 12101876	1045	0.0
CRA 154000034710284	gi 10331724	1031	0.0
CRA 58000099305099	gi 12895306	1025	0.0

FIGURE 1B

Docket No.: CL001313-DIV
 Serial No.: (to be assigned)
 Inventors: Chunhua YAN et al.
 Title: ISOLATED HUMAN KINASE...

CRA 222000001550936	gi 15446747	1021		0.0
CRA 78000105667113	gi 10352757	1017	0.0	
CRA 3000001081732	gi 2877701	1009		0.0
CRA 63000075437648	gi 14804614	1003	0.0	
CRA 114000006533071	gi 8142744	985		0.0
CRA 149000089175289	gi 13402435	981		0.0
CRA 155000041341862	gi 10144531	977		0.0
CRA 1000490895790	gi 5433043	954		0.0

EXPRESSION INFORMATION FOR MODULATORY USE:

library source:

gi Number	Organ	Tissue Type
gi 13994615	brain	hypothalamus
gi 10999153	PLACE1	placenta
gi 10947466	MAMMA1	mammary gland
gi 10996930	PLACE1	placenta
gi 15438670	brain	hippocampus

FIGURE 1C

1 MGEMQGALAR ARLESLLRPR HKKRAEAQKR SESFLLSGLA FMKQRRMGLN
51 DFIQKIANN S YACKHPEVQS ILKISQPQEP ELMNANPSPP PPSQQINLG
101 PSSNPHAKPS DFHFLKVIGK GSFGKVLAR HKAEVVFYAV KVLQKKAILK
151 KEEKHIMSE RNVLLKNVKH PFLVGLHFSF QTADKLYFVL DYINGGELFY
201 HLQRERCFLE PRARFYAAEI ASALGYLHSL NIVYRDLKPE NILLDSQGHI
251 VLTDFGLCKE NIEHNSTTST FCGTPEYLAP EVLHKQPYDR TVDWWCLGAV
301 LYEMLYGLPP FYSRNTAEMY DNILNKPLQL KPNITNSARH LLEGLLQKDR
351 TKRLGAKDDF MEIKSHVFFS LINWDDLINK KITPPFNPNV SGPNDLRHFD
401 PEFTEEPVPN SIGKSPDSVL VTASVKEAAE AFLGFSYAPP TDSFL (SEQ ID NO:2)

FEATURES:

Functional domains and key regions:

Prosite results:

PDOC00001 PS00001 ASN_GLYCOSYLATION

N-glycosylation site

Number of matches: 4

1	58-61	NNSY
2	265-268	NSTT
3	333-336	NITN
4	389-392	NVSG

PDOC00004 PS00004 CAMP_PHOSPHO_SITE

CAMP- and cGMP-dependent protein kinase phosphorylation site
380-383 KKIT

PDOC00005 PS00005 PKC_PHOSPHO_SITE

Protein kinase C phosphorylation site

Number of matches: 4

1	159-161	SER
2	337-339	SAR
3	351-353	TKR
4	424-426	SVK

PDOC00006 PS00006 CK2_PHOSPHO_SITE

Casein kinase II phosphorylation site
424-427 SVKE

PDOC00007 PS00007 TYR_PHOSPHO_SITE

Tyrosine kinase phosphorylation site
130-138 RHKAEVVFY

FIGURE 2A

PDOC00008 PS00008 MYRISTYL
N-myristoylation site
175-180 GLHFSF

PDOC00100 PS00107 PROTEIN_KINASE_ATP
Protein kinases ATP-binding region signature
118-150 IGKGSFGKVLLARHKAEVIFYAVKVLQKKAILK

PDOC00100 PS00108 PROTEIN_KINASE_ST
Serine/Threonine protein kinases active-site signature
232-244 IVYRDLKPENILL

Membrane spanning structure and domains:

Helix	Begin	End	Score	Certainty
1	293	313	0.948	Putative

BLAST Alignment to Top Hit:

>CRA|67000082668077 gi|14756346 /def=ref|XP_037046.1|
(XM_037046) serum/glucocorticoid regulated kinase [Homo
sapiens] /org=Homo sapiens /taxon=9606 /div=PRI
/dataset=nraa /length=431
Length = 431

Score = 843 bits (2154), Expect = 0.0
Identities = 406/407 (99%), Positives = 407/407 (99%)
Frame = +1

Query: 292 LAFMKQRRMGLNDFIQKIANN SYACKHPEVQSILKISQPQEP ELMNANPSPPPSPSQQIN 471
+AFMKQRRMGLNDFIQKIANN SYACKHPEVQSILKISQPQEP ELMNANPSPPPSPSQQIN

Sbjct: 25 IAFMKQRRMGLNDFIQKIANN SYACKHPEVQSILKISQPQEP ELMNANPSPPPSPSQQIN 84

Query: 472 LGPSSNPHAKPSDFHFLKVI GKGSGFKVLLARHKAEVIFYAVKVLQKKAILKKKEEKHIM 651
LGPSSNPHAKPSDFHFLKVI GKGSGFKVLLARHKAEVIFYAVKVLQKKAILKKKEEKHIM

Sbjct: 85 LGPSSNPHAKPSDFHFLKVI GKGSGFKVLLARHKAEVIFYAVKVLQKKAILKKKEEKHIM 144

Query: 652 SERNVLLKNVKHPFLVGLHFSFQTADKLYFVL DYINGGELFYHLQRERCFL EPRARFYAA 831
SERNVLLKNVKHPFLVGLHFSFQTADKLYFVL DYINGGELFYHLQRERCFL EPRARFYAA

Sbjct: 145 SERNVLLKNVKHPFLVGLHFSFQTADKLYFVL DYINGGELFYHLQRERCFL EPRARFYAA 204

Query: 832 EIASALGYLHSLNIVYRDLKPENILLDSQGHIVLTDFGLCKENIEHNSTTSTFCGTPEYL 1011
EIASALGYLHSLNIVYRDLKPENILLDSQGHIVLTDFGLCKENIEHNSTTSTFCGTPEYL

Sbjct: 205 EIASALGYLHSLNIVYRDLKPENILLDSQGHIVLTDFGLCKENIEHNSTTSTFCGTPEYL 264

FIGURE 2B

Query: 1012 APEVLHKQPYDRTVDWWCLGAVLYEMLYGLPPFYSRNTAEMYDNILNKPLQLKPNITNSA 1191
APEVLHKQPYDRTVDWWCLGAVLYEMLYGLPPFYSRNTAEMYDNILNKPLQLKPNITNSA
Sbjct: 265 APEVLHKQPYDRTVDWWCLGAVLYEMLYGLPPFYSRNTAEMYDNILNKPLQLKPNITNSA 324

Query: 1192 RHLLEGLLQKDR TKRLGAKDDFMEIKSHVFFSLINWDDLINKKITPPFNPVSGPNDLRH 1371
RHLLEGLLQKDR TKRLGAKDDFMEIKSHVFFSLINWDDLINKKITPPFNPVSGPNDLRH
Sbjct: 325 RHLLEGLLQKDR TKRLGAKDDFMEIKSHVFFSLINWDDLINKKITPPFNPVSGPNDLRH 384

Query: 1372 FDPEFTEEPVPNSIGKSPDSVLVTASVKEAAEAFLGFSYAPPTDSFL 1512
FDPEFTEEPVPNSIGKSPDSVLVTASVKEAAEAFLGFSYAPPTDSFL
Sbjct: 385 FDPEFTEEPVPNSIGKSPDSVLVTASVKEAAEAFLGFSYAPPTDSFL 431 (SEQ ID NO:4)

Hammer search results (Pfam):

Scores for sequence family classification (score includes all domains):

Model	Description	Score	E-value	N
PF00069	Eukaryotic protein kinase domain	298.1	1.1e-85	1
PF00433	Protein kinase C terminal domain	56.0	5.6e-16	1
CE00022	CE00022 MAGUK_subfamily_d	24.7	3.5e-07	1
CE00359	E00359 bone_morphogenetic_protein_receptor	14.6	0.0019	2
PF00787	PX domain	7.2	2.4	1
CE00031	CE00031 VEGFR	0.6	2.7	1
CE00292	CE00292 PTK_membrane_span	-49.8	3.8e-06	1
CE00287	CE00287 PTK_Eph_orphan_receptor	-53.1	7.6e-05	1
CE00289	CE00289 PTK_PDGF_receptor	-67.6	0.37	1
CE00291	CE00291 PTK_fgf_receptor	-87.2	0.00097	1
CE00286	E00286 PTK_EGF_receptor	-88.2	1e-05	1
CE00290	CE00290 PTK_Trk_family	-153.3	9.1e-05	1
CE00016	CE00016 GSK_glycogen_synthase_kinase	-159.4	8.3e-08	1

Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
PF00787	1/1	40	75 ..	108	147 .]	7.2	2.4
CE00359	1/2	115	143 ..	144	175 ..	0.5	19
CE00289	1/1	110	212 ..	1	109 []	-67.6	0.37
CE00031	1/1	216	260 ..	1051	1095 ..	0.6	2.7
CE00359	2/2	232	283 ..	272	327 ..	12.8	0.0064
CE00290	1/1	112	341 ..	1	282 []	-153.3	9.1e-05
CE00291	1/1	112	345 ..	1	285 []	-87.2	0.00097
CE00286	1/1	112	349 ..	1	263 []	-88.2	1e-05
CE00292	1/1	112	354 ..	1	288 []	-49.8	3.8e-06
CE00022	1/1	223	357 ..	133	275 ..	24.7	3.5e-07
CE00287	1/1	112	367 ..	1	260 []	-53.1	7.6e-05
PF00069	1/1	112	369 ..	1	278 []	298.1	1.1e-85
CE00016	1/1	41	442 ..	1	433 []	-159.4	8.3e-08
PF00433	1/1	370	444 ..	1	70 []	56.0	5.6e-16

FIGURE 2C

```

1 TCTGGCTCGT GCTCTCATGT CATCTCAGAG TTCCAGCTTA TCAGAGGCAT
51 GTAGCAGGGA GGCTTATTCC AGCCATAACT GGGCTCTACC TCCAGCCTCC
101 AGAAGTAATC CCCAACCTGC ATATCCTTGG GCAACCCGAA GAATGAAAGA
151 AGAAGCTATA AAACCCCTT TGAAAGGTTT GTACTTACCG TACTATATTT
201 TGCAGATGCC TCAAAGGATT TGGGGTTACT TGGCATGGGG AAGGCACATA
251 AGGTGGGGTG TAGGAGAGGG TCTCTGGTTG TAGGTTTCTT AATTTAATGT
301 TTGAAAACAA ACATGCAAAA GTCTGTGTGC AGGTTGATGT TTCTGGGCAG
351 CCTGAGCAAA ATTTGCTCTC TCAAGAGGGA AAGGAACCAG GTGGGAGCAG
401 AGCTAGGCTG GGCTAGGCTA GTTGAATGGT GGGACATGAC ATACGGGTGG
451 CACTGGCAAT AACAAAGTCA CATTCTATGA AGATTCCCTG CAAGAGGAAG
501 CAGACATGGG CCAGTTACTG TGATTGAAA TTGCCTAAAC ATTGCTTTAG
551 GTTGGCATGT CAATTTGAGG TACTAGTGTT TTTTGTGTT TTGTTTTGT
601 TTTGTTTTTG TTTGTTTGT TGTTTTGAGA CGGAGTCTCG CTCTGTTGCC
651 AGGCTGGAGT GCAGTGGCGT GATCTCGGCT CACTGCAACC TCCGCCTCCC
701 GGGTTCAAGC GATTCTCTG CCTCAGCCTC CCGAGTAAC TGGACTACAG
751 GCGCACGCCA CCACGCCTGG CTAATTTTTT TATTTTCAGT AGAGACGGGG
801 TTTACCATG TTAGCCAGGA TGGTCTCGAT CTCTTGACCT CGTGATCCGC
851 CCGCCTTGGC CTCCCAAAGT GCTGGGATTA CAGGCGTGAG CCACTGCGCC
901 CGGCCCCAGT AAATGCTTTT TATAAGTGTG GGCAGTGGC AAATTTTCCC
951 AGCCAGACTC CAGGAGAGAG AATGTGTTTC CTTCTCTCG GTTTGGGGCT
1001 GTTGCAACAA AGCAAACCAA GGAGTTGAGA CTAGAGCTCA CTTTAGGGCA
1051 AGTGGGGGTG GTTTTGCTG CAAAACAAAC CCCTGCCCAA GACCAAGGAA
1101 AAGGCGTTTC ACATGCTATT CCTGGTTTGA CAGCTGGTAT TTCGGGACTG
1151 TGCCAGATCC AGTAGGCAAC TTTAAAATGG CAGAGCCTTT GGTAGCAAGA
1201 GGTGATGGCA GGGCAGCCAC CGCAGACAGC AACAGCGAGC GCCAGGTACC
1251 TGGCCCTGCG AATAGTGGTA ACTTGTAAC TCCCGCTCCG GGCCAGTCCG
1301 CTGTGCTCGC GGCTTCCCGG CCAGCACTGG CTCACGTCCC CGCGCCGGCG
1351 GTCAGGCTGC GGCTCCAGA CATCCCCCAG CCGCGGGGT ACTGGAAGGC
1401 ACCGGCATCG CTGTTCTGCA GAGCCCGGGC CGCCGCTCG AGCTTCCCTC
1451 TCTTCCCTGC CTTCTGCAGC GGAGTACCC GGCTAATCTT TCAGGATAAA
1501 GTCACAGTTT ATGTGGGACT CACATAAAGA GCGAGCGAGG TGGCAAACT
1551 AAGAAGCCCT GGGGAGCCT TGAGTTAAAC CCAGGGAGGG TAGGGACGAT
1601 TTTAAGACCA TGTATCATGA CCTGCAGGGT TTTAGGTGG GACAGCGGGA
1651 GAGGAGCAGG CCCACAGAG GAATCGAGGA TGCCCGGTT ACGCCAGGTC
1701 TGCCCCCGGG CAAAGCTACC CCTCCCTTCG CTTGTTACCT CCTCACGTGT
1751 TCTTGGCATG GCAGAGATTA AAAATGCAAG GAAAAAATT ACATGCGGAA
1801 CGGACAAAAT GTTCTCAGAG ATTACTTCAG AAAAAAAAAA GTGAAATGCA
1851 GATTGTACTT CTTCTTTAG TGCAGAGACG ACTTTTATTT CCGCCCCCTC
1901 CCTCCACAT TCCTGACCTC TCCCTCCCC TTTTCCCTCT TTCTTTCCTT
1951 CCTTCTCCT CTTCCAAGT CTGGGATTTT TCAGCCTTGC TTGGTTTTGG
2001 CAAAAGCAC AAAAAAGGCG TTTTCGGAAG CGACCCGACC GTGCACAAGG
2051 GCCATTTGTT TGTTTTGGGA CTCGGGGCAG GAAATCTTGC CCGGCCTGAG
2101 TCACGGCGGC TCCTTCAAGG AAACGTCAGT GCTCGCCGGT CGCTCTCGTC
2151 TGCCGCGCGC CCCGCCGCC GCTGCCCATG GGGGAGATGC AGGGCGCGCT
2201 GGCCAGAGCC CGGCTCGAGT CCCTGCTGCG GCCCCGCCAC AAAAAGAGGG
2251 CCGAGGCGCA GAAAAGGAGC GAGTCCTTCC TGCTGAGCGG ACTGGGTAAG
  
```

FIGURE 3A

2301 CGCCGCCGCC GGGCCGCTG GGGGCTTGGC TCACTTCCCC AGAGCGGCTT
2351 GGAGGCAGGG GCCGGCTTTC GTCGGAGTTC TCGGGGCCGG GGTCCCCGGC
2401 GCGGGAACGG GAGGACCTGG CGGGCGAGGT CGCGCGCGCA GGCCTGCGCC
2451 CCAGGGATAA ACCCCGGAGG GTGGCGCGCA CCGCCGGCTC GGGTTGGGGA
2501 GGAGGGTGGG AGTCCGGCCG CAGGACGGCG CCTGGCCGGG GAGAGGGTAT
2551 CTGCAGGGAC AGTGAGCGAA GCCACCGTGG CCGCCGCGCA CCCGCCGGGA
2601 AGCGCTTCGG CGCTGCGAAC CCGGCTTTCT CCGGCGGCGG AATAAATGAG
2651 AGAGGTGGAA AACTACCCCG GGCTCTCCGG CCCTCCCCGC GCCCTCCGCC
2701 GGCGGTTCT CTCTCTCTG CCCAGGAGC CGATGGAGAC TGATAACGGC
2751 CCTGCGCCAG GCCGTCCCCG GCGGTCCCTC GCGCCCCCGC CCGGGGCTCG
2801 CCCTCTCAAT GGGGACAGAA CCGCCCGCCG CAGGCAGCGT AGCCGCCAGC
2851 AAACCGCGAG GCGGTGCGGG CGGGCGGAGG GGCGAGGCGA AGGGCGGGGC
2901 CACTTCTCAC TGTCGCGCAG GCCCCGCCCC CGCGGCGGTG CCTTTTTTAT
2951 AAGGCCGAGC GCGCGGCCTG GCGCAGCATA CGCCGAGCCG GTCTTTGAGC
3001 GCTAACGTCT TTCTGTCTCC CCGCGGTGGT GATGACGGTG AAAACTGAGG
3051 CTGCTAAGGG CACCCTCACT TACTCCAGGA TGAGGGGCGAT GGTGGCAATT
3101 CTCATCGGTG AGTGCAGGAA TCTTGCGGGA CTCTGCTCC AGGAGACGCA
3151 AAGTGGAAAT TTTTGAAG TCCCGATCA GATTAGTGTG TGTGGCGCCG
3201 GACGTTATGA AGCCGTCTAA ACGTTTCTTT ATTCTCCTC CTTCATCCAC
3251 AGCTTTCATG AAGCAGAGGA GGATGGGTCT GAACGACTTT ATTCAAGA
3301 TTGCCAATAA CTCCTATGCA TGCAAACAGT AAGTTTGACC GGATTTGAGG
3351 AAATAACTAG TATAGTTTGA ATTTGCCAGC GGTAACATT CTCATCACGG
3401 CGTTTATCGG GAAGGCGAAG ACTTCTTCTG GGTGGGGAT CTCATTTCTC
3451 CTAAATTCT AATATATTG ACACATTTTA AACATTAAAG TTAATTGCT
3501 GATTTGGCTT GAACTGGAGA TGTAAGATAA ATGGTTCGTG TTGGCCGAAT
3551 TCACGGCCTT TCTCCATGAG CAACAATCCT TATTTCTGTA TTTAATGGGG
3601 TTTATTATTT TCTTAACTG ACTAATGTAT TGGGGTATTT TCAGTTTAAA
3651 CAGTGAATTA TCCGGGTAGA AGTCGGTAGA GCCAGGAAAC TCACTTTTGA
3701 TGTTGGTGTG CCCCCTAGTG GCGAGCTGGA TTCTAAATCG TGCCCTTTAT
3751 TCCCTGCAGC CCTGAAGTTC AGTCCATCTT GAAGATCTCC CAACCTCAGG
3801 AGCCTGAGCT TATGAATGCC AACCCTTCTC CTCCAGTAAG TTTTGTATG
3851 TGCCGTGCAT CTGTGGAGAA CTGTAAGGGA GTCAGTTAGT ATTCCTACAT
3901 TAATGGATTA AAATAGCATT TCTAGAAATT AGTATCAAGG CAGGAATGCT
3951 TCATTATGGC ATAACAAGTG ATATAAATAT TTAAGTATTG AGTCAGAGTA
4001 TTATTTTATT TTTTCTCTG GCATATTTTA CCTCCAAAGT GGTATTTTA
4051 AAAGGCATAT TTCATAAAAA GGTTTTATCT GTCTGAAACA ACATGACTGT
4101 GTGCAGTTTC CATACTCATT TGAAATGTGA TGAAATGTAG TTTTGAATGT
4151 TTATAGATGT ATGGTCATTT GCATCAGTCA TTTGTAGATG TAACATTTTC
4201 TACATCGTTT ATGTTATAGA TGTCTTCTT TGAAGCAATG GTATTTAAAG
4251 AAATTCCTTT TTTTTTTTTC TAGCCAAGTC CTTCTCAGCA AATCAACCTT
4301 GGCCCGTCGT CCAATCCTCA TGCTAAACCA TCTGACTTTC ACTTCTTGAA
4351 AGTGATCGGA AAGGGCAGTT TTGGAAAGGT AATTTCAAAT CTGAAGATCT
4401 TTTGGTACAC TTCCTTCATG TCCTCTTTTA TATTCTCCCT GGATGAGGAT
4451 AGAAAAATGA TTTTTTTAAA TTGAAATTTT AGGTTCTTCT AGCAAGACAC
4501 AAGGCAGAAG AAGTGTTCTA TGCAGTCAAA GTTTTACAGA AGAAAGCAAT
4551 CCTGAAAAAG AAAGAGGTAT GAGATGTGCT TGATGGGGCT GGCATTGGCG

FIGURE 3B


```

4601 GTAGACACTC CTTGAATAAT CTTGATTCTG GAATGTTGGT GCCAAGTTGA
4651 AACATGCCAC TAAATCTGAA TCGTCATTTT CCTAGGAGAA GCATATTATG
4701 TCGGAGCGGA ATGTTCTGTT GAAGAATGTG AAGCACCTT TCCTGGTGGG
4751 CCTTCACTTC TCTTTCCAGA CTGCTGACAA ATTGTACTTT GTCCTAGACT
4801 ACATTAATGG TGGAGAGGTG AGCAGGGGGG ATAGAAGTCA ACTCTTAGTG
4851 TCTCTGCACA GCCTGCTTTG TTTTAGTTTG AGAAAAAAGT TTTCAAAGAT
4901 TTTTGGTGGG GAGAATGTTA CCAGAATTAG CATTTCTTC AACCTGTCAG
4951 GTTTATAGTT AATAGATTAC TTGGGGCCAC TTCCTGCAGT TGTTCTTTTG
5001 CTGTGTATGT CAAAACTAAT TAAATTCATT TGCAACCCAG AATGACTTTG
5051 TTCTGTCTCC TGCAGTTGTT CTACCATCTC CAGAGGGAAC GCTGCTTCCT
5101 GGAACCACGG GCTCGTTTCT ATGCTGCTGA AATAGCCAGT GCCTTGGGCT
5151 ACCTGCATTC ACTGAACATC GTTTATAGGT AAGCCTGAGA GCTCTTCAGG
5201 CTACCAGTTT TGGTATAAAG GAGACGTAGC ACTGGCTGTT TCATAGGGCC
5251 TTAAAATAAT TTGTGTTTAT TTGCAACTTG GTTGCCTAAA ACCAGATCCC
5301 CTAGCACGTG AGCTGGCTTG ACTTAAGTGC CAAGGGGGAA CCAGCCAAGT
5351 AGGATTGTGC CTAATCCAGA ATAGATGAGC AGAACAAGGG CTCCCTTTTT
5401 TCTTCACTAC ACAACTACAG TGAACCTAAA ATGCCTCTAA TACCTTTAGC
5451 AATTATCTTT AAGAGGATAT CTTATGAAGT GAAATTAAGT TGTGCAACTA
5501 CTTTTCTATT CACTTTTTTA CAGAGACTTA AAACCAGAGA ATATTTTGCT
5551 AGATTCACAG GGACACATTG TCCTTACTGA CTTGCGACTC TGCAAGGAGA
5601 ACATTGAACA CAACAGCACA ACATCCACCT TCTGTGGCAC GCCGGAGGTA
5651 GGCGCTGTCT TGGTTTGGTG CCTGGTTTAC CCCCCTTC CAAGAGAGAG
5701 ATGTACAATC ATGCACTTAA CTACCAAAAA GAGTAAACTC CTCTCAGAGA
5751 CTTCTTAATA CAGTTCAGTG CAAATAAAAT ACATTTGCTG TTTGATGTAG
5801 CATGAGAAAT CCAAGTCCT TCTGTTCTT TACTGAAAAG TAGCTGTTTG
5851 TAAGTAAGAT CTGCATCATA AAAACTTTCT AAATCCCTAA GTAAGAGATA
5901 TCAAGTGCCC AGCAGTTTCC TAAATGTCAG TACACATAGG TAGCCAGTCA
5951 CCCTCAAAAA GTCCAGCAGT TTTATCAGGA AGGAATCTAA AGATATCTAT
6001 CTTCCAAGCT GGCTCTGGGT CTCTCAGCTT TTTCAAACTA AATGTGTGGT
6051 CGTGGGATTG CTTGCTTTCG CAGGTTCTAA ACGCTGTTTC CCTGGTCTGT
6101 TTTTCAGTAT CTCGCACCTG AGGTGCTTCA TAAGCAGCCT TATGACAGGA
6151 CTGTGGACTG GTGGTGCTG GGAGCTGTCT TGTATGAGAT GCTGTATGGC
6201 CTGGTGAGTG GCACATTGGG AACCATGGAA CACTGCCTGC TCCCTACAAT
6251 ATTGCCTTCA CACAGCCCAT GCTTGGCCAT GGTGTCTTGC CTTACCAGT
6301 ACGCTTATCA AAAGCAGCTA AGAGGCATAT TGGTTATTTT ATAGTTCATA
6351 AGAATAATCA CTTACCTGGT TCTTTTGTGC ATTTACATT TTAGTAGATA
6401 GGACCACATT GAACCTGTGT GGTGGTGAAA AACTACCACT TATTAACATC
6451 TACCCCTCA CCCTCCACAC ACACACACAC AAACACACAC ACGGGTTGCA
6501 AAGTAGACAC TTAAATAGCA AGGGAAAAGA AAGCATTGAG GTGGGGAGAG
6551 TTTCTCAAAT CGAGCCTAAT ATTTATTGCC GTTTATATCT TTTTCTCTAC
6601 TGGTAATGTG TGCCATATGA AACTCCAAT TAAGTCTAAA GTAATTTTCC
6651 CCTTCTTTCA GCCGCTTTT TATAGCCGAA ACACAGCTGA AATGTACGAC
6701 AACATTCTGA ACAAGCCTCT CCAGCTGAAA CCAAATATTA CAAATCCGC
6751 AAGACACCTC CTGGAGGGCC TCCTGCAGAA GGACAGGACA AAGCGGCTCG
6801 GGGCCAAGGA TGACTTCGTG AGTGATGTTT TCCTGTCCTC CTGGGCCGGC
6851 CGGGACGTGC ACTAGACCTC CCTGCCCTTA TTGAATGCAC CTGTCTAAAT

```

FIGURE 3C

```

6901 TAATCTTGGG TTTCTTATCA ACAGATGGAG ATTAAGAGTC ATGTCTTCTT
6951 CTCCTTAATT AACTGGGATG ATCTCATTAA TAAGAAGATT ACTCCCCCTT
7001 TTAACCCAAA TGTGGTGAGT ATCTGTCTCT CTTCTAAGTA TAGAGAAGCC
7051 CAAAGGGCAT TTATTTTAAT TCAGAATTGT CTGGGGGAGG GTTGGGAAGGA
7101 ATACATTGGC AGATGTTTTT TCCATAAACC TGTTATTTTA CCTACATAAA
7151 AAGCACATTT TTGTGTCCCA ACAAGGCTCC CATAATTTTT AGACACATTT
7201 ATCAATTCGA AGCACCAAAA GGCAACAAGT GAACATTATT CTTATGTTTA
7251 ACTGTGTGTA GCCTTTTGAG ATTTTGTGCT TGAAGTGGGT GATTATGGAA
7301 GTTGATATAA GACTTAAACT TGGTATTTAA AGCCTGGTCA AGATTTCCTT
7351 GTCCTGTGTC TAGTGTGAGT TCTTGACAAG AGTGTTTTTC CCTTCCCGTC
7401 ACAGAGTGGG CCCAACGACC TACGGCACTT TGACCCCGAG TTTACCGAAG
7451 AGCCTGTCCC CAACTCCATT GGCAAGTCCC CTGACAGCGT CCTCGTCACA
7501 GCCAGCGTCA AGGAAGCTGC CGAGGCTTTC CTAGGCTTTT CCTATGCGCC
7551 TCCCACGGAC TCTTTCCTCT GAACCCTGTT AGGGCTTGGT TTTAAAGGAT
7601 TTTATGTGTG TTTCCGAATG TTTTAGTTAG CTTTTGGTG GAGCCGCCAG
7651 CTGACAGGAC ATCTTACAAG AGAATTTGCA CATCTCTGGA AGCTTAGCAA
7701 TCTTATTGCA CACTGTTTCG TGGAAAGCTT TTGAAGAGCA CATTCTCCTC
7751 AGTGAGCTCA TGAGGTTTTC ATTTTATTC ATCGTCCAA CGTGGTGCTA
7801 TCTCTGAAAC GAGCGTTAGA GTGCCGCTT AGACGGAGGC AGGAGTTTCG
7851 TTAGAAAGCG GACGCTGTTT TAAAAAAGGT CTCCTGCAGA TCTGTCTGGG
7901 CTGTGATGAC GAATATTATG AAATGTGCCT TTTCTGAAGA GATTGTGTTA
7951 GCTCCAAAGC TTTTCCTATC GCAGTGTTC AGTTCTTTAT TTTCCCTTGT
8001 GGATATGCTG TGTGAACCGT CGTGTGAGTG TGGTATGCCT GATCACAGAT
8051 GGATTTTGTG ATAAGCATCA ATGTGACACT TGCAGGACAC TACAACGTGG
8101 GACATTGTTT GTTCTTCCA TATTTGGAAG ATAAATTTAT GTGTAGACTT
8151 TTTTGTAAGA TACGGTTAAT AACTAAAATT TATTGAAATG GTCTTGCAAT
8201 GACTCGTATT CAGATGCTTA AAGAAAGCAT TGCTGCTACA AATATTTCTA
8251 TTTTLAGAAA GGGTTTTTAT GGACCAATGC CCCAGTTGTC AGTCAGAGCC
8301 GTTGGTGTTT TTCATTGTTT AAAATGTCAC CTGTAAAATG GGCATTATTT
8351 ATGTTTTTTT TTTTGCATTC CTGATAATTG TATGTATTGT ATAAAGAACG
8401 TCTGTACATT GGGTTATAAC ACTAGTATAT TTAACTTAC AGGCTTATTT
8451 GTAATGTAAA CCACCATTTT AATGTACTGT AATTAACATG GTTATAATAC
8501 GTACAATCCT TCCCTCATCC CATCACACAA CTTTTTTTGT GTGTGATAAA
8551 CTGATTTTGG TTTGCAATAA AACCTTGAAA AATATTTACA TATATTGTGT
8601 CATGTGTTAT TTTGTATATT TTGGTTAAGG GGGTAATCAT GGGTTAGTTT
8651 AAAATTGAAA ACCATGAAAA TCCTGCTGTA ATTTCTGCT TAGTGGTTTG
8701 CTCCCAACAG CAGTGGTTTC TGACTCCAGG GGAGTATAGG ATGGTCTTAA
8751 AGCCAACCTA CGTTCAGGC CTTTTAGCA GCATTTTATG GTGTCTGTCA
8801 TTCATAAATC CATCCAAGGA AATCCTTGC AATTTACTCA TCTTGCAAGG
8851 ATTGCTATGA AGTAATGCTT CCTGTATTTA TTGCCTGTCC TGTGAAGTTG
8901 GACTATTTGT CCTGACATTT GGCTTGCTT CAGTTACAGG TAATTCCTTC
8951 CAGAAATATT TGAAAGCCTA CTCTGGGCTC TATTGCGAGT GCTCAGGATA
9001 TCGTAGTGGA CAAAGCAGAC AACTTCGCCC TTCCAGAGCC TGATGAAGAA
9051 GGCCGACCTA AAGCAGTTAG TTGAGATGGA AATTGAGAAA TAGTCTGTGA
9101 AGTTTAGGAG AATGCCACAC AAGAGGGTGA GAATTTTTTT TTTTTTTTTT
9151 TTTTTTTTGG AGACACGGTC TTA CTGTGTC GCCCAGGCTG GAGTGCAGTG

```

FIGURE 3D

```

9201 GTGTGATCTT GGTTCAGTGC AGCCTCCGCC TCCTGGGTTC ATGTGATCCC
9251 CCCATCTCAG CCTCCTGAGT AGCTGGGACT ACAGGCATGC ACCACCATGC
9301 CTGGCTAATT TTTGTATTTT AGTAGAGATG GGATTTTACC ATGTGGGCCA
9351 GGCTGGTCTC GAATCCCTGG CCTCAAGTGA TCTGTCTGCC TCGGCTTCCC
9401 TAAGTGCTGG GGAGAATGTT TAAATAAGT GGATATGTTC CAAAAAGCT
9451 GACCTGGCTG GGACATCTGG TTTCTGAGAG TACCTGGAGT TGACCCAGGT
9501 CTAGAGTGAG CTCAGTAAAG GGACCCTGAA GGAGCTCATC CCTAGCTTGG
9551 ACTGAAGCTT CTTGAGCCAG TGTCTACCTA GCACCCTAAG GGCCCAGCAG
9601 GCTCTGGGGC TGTGTGGCAG AGCCCACTCC TAGAGCTCAC CCCACTGTGA
9651 TATTACCTGT GGGAGAAAGC GAGGTGGCAC CATCCTTGGA GATCTTGAGT
9701 CCAAAGGTTT GGACTTTTTT ACTCTTCTAG GCCTTCCACA CAAATACTTA
9751 ACAAATAATC AGGGAATCCC CAAACAGTTG ATGTTGCTGC TGCCTTAATT
9801 GCAAAAGCAC CCTGTAGGCC TGCTGCACCC CCGCTACCTT GACCTTCCAG
9851 TTCGCACAGG GATTTCCCCA AGGGAAAGCT GTGAGCTTTT TTCCTCTTAT
9901 CCTTGCTCTT GGGTCTCACC TCACTTTGCC TCAGTCCCCC TCTCTACCC
9951 CACAAGGTTT CCAAGGGCCA AACAGGTGTT CAGAGATAAC CGAGTTCCTC
10001 TCCCTCATGA TCTAATGAAG GAAGAAGATG AAAACGAGTC GATAGCTTTT
10051 TGCTCAAGGT GGGCCACCGG TCATGCTCTG CTGTTGACTT ACTGCTCTAC
10101 AGGCATTAGC TACGTGTTCA ATTCCCTACC GGGCCAGTTT GACAAATAAA
10151 GAGTCCAAAG CAAGGCCAGG CACGGTGGCT CACGCTTGTA ATCCCAGCAC
10201 TTTGGGAGGC CGAGGCGGGC AGATCACGAG GTCAGGAGAT CGAGACCATC
10251 CTGGCTAACA TGGTGAAACC CCGTCTCTAC TAAAAATACA AAAAAATTAG
10301 CCGGGCGTGG TGGTGGGCGC CTGTAGTCCC AGCTACTCGG GAGGCTGAGG
10351 CAGGAGAATG GCGTGAACCA GGGAGGCGGA GCTTGCAGTG AGCCGAGATC
10401 GCACCACTGC ACTCCAGCCT GGGCGACAGA GCAAGACTCT GTCTCAAAAA
10451 ACAAACAAA ACAAAGCAT GTATTTTCTT ATTAAGATT GATGCCGGCT
10501 CTAACATAGA GACTCATTGC ATATTCCCC TCATTCTCAT TCTCAATAAC
10551 AGTTATGAAT TCCTCTCGA ACA (SEQ ID NO:3)

```

FEATURES:

Genewise results:

```

Start: 2178
Exon: 2178-2295
Exon: 3253-3328
Exon: 3760-3835
Exon: 4274-4378
Exon: 4483-4566
Exon: 4686-4817
Exon: 5066-5178
Exon: 5524-5647
Exon: 6108-6203
Exon: 6662-6817
Exon: 6925-7014
Exon: 7405-7569
Stop: 7570

```

FIGURE 3E

Sim4 results:

Exon: 2178-2295, (Transcript Position: 1-118)
 Exon: 3253-3328, (Transcript Position: 119-194)
 Exon: 3760-3835, (Transcript Position: 195-270)
 Exon: 4274-4378, (Transcript Position: 271-375)
 Exon: 4483-4566, (Transcript Position: 376-459)
 Exon: 4686-4817, (Transcript Position: 460-591)
 Exon: 5066-5178, (Transcript Position: 592-704)
 Exon: 5524-5647, (Transcript Position: 705-828)
 Exon: 6108-6203, (Transcript Position: 829-924)
 Exon: 6662-6817, (Transcript Position: 925-1080)
 Exon: 6925-7014, (Transcript Position: 1081-1170)
 Exon: 7405-7572, (Transcript Position: 1171-1338)

CHROMOSOME MAP POSITION:

chromosome 6

ALLELIC VARIANTS (SNPs):

DNA

Position	Major	Minor	Domain
738	A	G	Intron

Context:

DNA

Position

738 GACATACGGGTGGCACTGGCAATAACAAAGTCACATTCTATGAAGATTCCCTGCAAGAGG
 AAGCAGACATGGGCCAGTTACTGTGATTTGAAATTGCCTAAACATTGCTTTAGGTTGGCA
 TGTC AATTTTCAGGTA TAGTGT TTTT TTTT GTTTT GTTTT GTTTT GTTTT GTTTT GTTT
 GTTT GTTTT GAGACGGAGTCTCGCTCTGTTGCCAGGCTGGAGTGCAGTGGCGTGATCTCG
 GCTCACTGCAACCTCCGCCTCCCGGGTTCAAGCGATTCTCCTGCCTCAGCCTCCCGAGTA
 [A,G]
 CTGGGACTACAGGCGCACGCCACCGCCTGGCTAATTTTCTATTTTCAGTAGAGACGG
 GGTTTCACCATGTTAGCCAGGATGGTCTCGATCTCTTGACCTCGTGATCCGCCCCGCTTG
 GCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACTGCGCCCGGCCCCAGTAAATGCTT
 TTTATAAGTGTGGGCACTGAGCAAAC TTTCCAGCCAGACTCCAGGAGAGAGAATGTGTT
 TCCCTTCTCTCGGTTTGGGGCTGTTGCAACAAAGCAAACCAAGGAGTTGAGACTAGAGCT

FIGURE 3F